

Please type a plus (+) inside this box - +

PTO/SB/29 (12/97)

Approved for use through 09/30/00. OMB 0651-0032

Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

<div style="writing-mode: vertical-rl; transform: rotate(180deg);"> 08/31/99  0595 U.S. PAT. FORM 37 </div>	<h2 style="margin: 0;">UTILITY PATENT APPLICATION TRANSMITTAL</h2> <p style="font-size: small; margin: 5px 0;">(only for nonprovisional applications under 37 CFR 1.53(b))</p>	Attorney Docket No. <span style="border: 1px solid black; padding: 2px;">9392-0011-999</span>	Total Pages <span style="border: 1px solid black; padding: 2px;">16</span>
	First Named Inventor or Application Identifier		
	Issac SAIAS		
	Express Mail Label No.		

<b>APPLICATION ELEMENTS</b> See MPEP chapter 600 concerning utility patent application contents.	<b>ADDRESS TO:</b> Assistant Commissioner for Patents Box Patent Application Washington, DC 20231
-----------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------

1. <input checked="" type="checkbox"/> Fee Transmittal Form <i>Submit an original, and a duplicate for fee processing</i> 2. <input checked="" type="checkbox"/> Specification <span style="float: right;">[Total Pages <u>14</u>]</span> <i>(preferred arrangement set forth below)</i> -Descriptive title of the Invention -Cross Reference to Related Applications -Statement Regarding Fed sponsored R&D -Reference to Microfiche Appendix -Background of the Invention -Brief Summary of the Invention -Brief Description of the Drawings <i>(if filed)</i> -Detailed Description of the Invention <i>(including drawings, if filed)</i> -Claim(s) -Abstract of the Disclosure <input checked="" type="checkbox"/> Drawing(s) <i>(35 USC 113)</i> <span style="float: right;">[Total Sheets <u>2</u>]</span> <input checked="" type="checkbox"/> Oath or Declaration <i>(unexecuted)</i> <span style="float: right;">[Total Sheets <u>2</u>]</span> a. <input type="checkbox"/> Newly executed (original or copy) b. <input type="checkbox"/> Copy from a prior application <i>(37 CFR 1.63(d))</i> <i>(for continuation/divisional with Box 17 completed)</i> <div style="text-align: center; font-weight: bold; font-size: small;">[Note Box 5 below]</div> i. <input type="checkbox"/> <b>DELETION OF INVENTOR(S)</b> Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33 (b). <input type="checkbox"/> Incorporation By Reference <i>(useable if Box 4b is checked)</i> The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.	6. <input type="checkbox"/> Microfiche Computer Program <i>(Appendix)</i> 7. <input type="checkbox"/> Nucleotide and/or Amino Acid Sequence Submission <i>(if applicable, all necessary)</i> a. <input type="checkbox"/> Computer Readable Copy b. <input type="checkbox"/> Paper Copy <i>(identical to computer copy)</i> c. <input type="checkbox"/> Statement verifying identity of above copies
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>ACCOMPANYING APPLICATION PARTS</b>	
8. <input type="checkbox"/> Assignment Papers (cover sheet & document(s)) 9. <input type="checkbox"/> 37 CFR 3.73(b) Statement <input type="checkbox"/> Power of Attorney <i>(when there is an assignee)</i> 10. <input type="checkbox"/> English Translation Document <i>(if applicable)</i> 11. <input type="checkbox"/> Information Disclosure <input type="checkbox"/> Copies of IDS Statement (IDS)/PTO-1449 Citations 12. <input type="checkbox"/> Preliminary Amendment 13. <input checked="" type="checkbox"/> Return Receipt Postcard (MPEP 503) <i>(Should be specifically itemized)</i> 14. <input type="checkbox"/> Small Entity <input type="checkbox"/> Statement filed in prior application, Statement(s) Status still proper and desired 15. <input type="checkbox"/> Certified Copy of Priority Document(s) and Transmittal Letter <i>(if foreign priority is claimed)</i> 16. <input type="checkbox"/> Other:	

17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information: <input type="checkbox"/> Continuation <input type="checkbox"/> Divisional <input type="checkbox"/> Continuation-in-part (CIP) of prior application No: filed .	
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

<b>18. CORRESPONDENCE ADDRESS</b>	
-----------------------------------	--

<input checked="" type="checkbox"/> Customer Number or Bar Code Label <span style="float: right;">20582</span> <div style="border: 1px solid black; padding: 2px; width: fit-content; margin: 5px auto;"> <i>(Insert Customer No. or Attach bar code label here)</i> </div>	<input type="checkbox"/> Correspondence address below
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------

NAME			
ADDRESS			
CITY	STATE	ZIP CODE	
COUNTRY	TELEPHONE	FAX	

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

# A METHOD FOR PERFORMING SOCIAL COMPUTATION

## FIELD OF THE INVENTION

The present invention relates generally to methods for performing social computation. More specifically, the present invention detects emergent concepts from a plurality of sites by creating an adjacency matrix representing the connectivity among the sites, computing the transpose of the adjacency matrix and computing the  $n$ th order eigenvalues of the product of the adjacency matrix and the transpose matrix.

## BACKGROUND

The main aim in "social computation" is to develop tools enabling the forecasting of the future behavior of a society. Most approaches in forecasting proceed in the following three-stepped approach. One postulates a parameterized dynamics of the underlying system. One then optimizes the choice of parameters to determine these parameters accounting best for the past observations. Finally one uses the calibrated dynamics to forecast future events.

For instance, most predictions done in the business community are based on statistical regressions. In that context, one postulates that the observable  $y$  is generated through a process  $y = f_{\lambda}(x) + \text{noise}$ :  $x$  is the explanatory variable, "noise" is a process with known dynamics and  $\lambda$  is a parameter to be calibrated. Linear regression corresponds to the assumption that "noise" is white, and to the choice of linear functions  $f_{\lambda}$ . The white noise assumption leads to mean-square optimization of the parameter: the past  $X_0$  yields a parameter  $\lambda_0$  all owing the prediction  $f_{\lambda_0}(x)$ .

Standard mathematical dynamical systems also proceed along this three-step approach, as do the modern agent-based models. Even though very different, all these forecasting methods rely on "proper" modeling of the underlying system dynamics. Most systems exhibit a chaotic behavior at small scales, so that only "skeletal models that tend to capture generic global dynamics and not microscale behavior" can hope to appropriately capture reality. Finer-scale prediction requires therefore a different approach.

The difference between detection and prediction is often just a matter of available technology. For example, until very recently, a pregnant woman had to await delivery to discover the gender of her child. Inferring this gender was therefore a predictive activity, trying to guess a fact that only future could reveal. Many people had argued that the only forecasting available was to flip a coin (with a small bias). The advent of new probing technology changed fully the paradigm of uncertainty. Now, uncertainty is not to be

dynamically revealed, (when one flips the coin, i.e., at delivery), but instead unveiled from a hitherto masked "random state". (Interestingly enough, the Turing model for random computation also assumes the existence of a hidden random tape consigning all future random flips.)

- 5 Many uncertain events are similarly not the product of dynamic random choices, but simply the emergence of facts so far kept "below the level of noise" for lack of appropriate technology. Many social phenomena fall within that level of uncertainty. Their so-called "unpredictability" is in fact more an expression of their complexity than of a genuine random or chaotic phenomenon of nature. The advent of the World Wide Web and
- 10 the emergence of new, dynamic, very large databases both raise new challenges and offer new possibilities for the acquisition of knowledge. On the one hand, their complexity seems to create new realms of uncertainty and unpredictability: conventional databases were "easy" to query and manipulate; but who can control the World Wide Web and the format of the displayed information? On the other hand, the new linkage of vast domains of
- 15 knowledge raises the possibility to investigate and corroborate facts that have been mostly disparate thus far.

Accordingly, there exists a need for a method for detecting emergent concepts from a plurality of sites.

20

## SUMMARY OF THE INVENTION

The present invention presents a method for partitioning that provides both a relevant metric and a set of clusters through an evolutionary learning process.

- It is an aspect of the present invention to present a method for detecting at least one emergent concept among a plurality of sites comprising the steps of:

- 25 creating at least one adjacency matrix  $A$ , said adjacency matrix having a plurality of entries,  $A_{i,j}$  wherein:

$i$  and  $j$  are among said plurality of sites;

$A_{i,j} = r$  if said sites,  $i, j$  are connected;

- 30  $A_{i,j} = 0$  otherwise; and

$r$  is a positive number;

computing the transpose matrix  $A^T$  of said adjacency matrix  $A$ ;

- computing the  $n$ th eigenvector  $X^{(n)}$  of a matrix product of said transpose matrix and said adjacency matrix,  $A^T A$  for determining an authority value of said plurality of sites, wherein  $n$  is a natural number.

35

It is an aspect of the present invention to present a method for detecting at least one emergent concept among a plurality of sites further comprising the steps of

computing the  $n$ th eigenvector  $\lambda^{(n)}$  of a matrix product of said adjacency matrix and said transpose matrix,  $A A^T$  for determining a hub value of said plurality of sites.

### BRIEF DESCRIPTION OF THE DRAWINGS

5           FIG. 1 provides a flow diagram of the method for detecting emergent concepts from a plurality of sites of the present invention.

FIG. 2 discloses a representative computer system in conjunction with which the embodiments of the present invention may be implemented.

10

15

20

25

30

35

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention presents methods for detecting emergent concepts from a plurality of sites. Without limitation, many of the following embodiments of these methods are explained in the illustrative contexts of the World Wide Web and intelligence applications. However, it will be apparent to persons of ordinary skill in the art that the aspects of the embodiments of the invention are also applicable in any context where emergent concepts can be detected from a plurality of sites.

The present invention is based on some very recent developments on the analysis of social linked structures as explained in Kleinberg J. (1998). *Authoritative Sources in a Hyperlinked Environment*. Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA-1998), ("Kleinberg") the contents of which are herein incorporated by reference. Social linked structures are further described in Gibson D., Kleinberg J. and Raghavan P. (1998), *Inferring Web Communities from Link Topology*. Proceedings of the 9<sup>th</sup> ACM Conference on Hypertext and Hypermedia, ("Gibson") the contents of which are herein incorporated by reference. In dynamic structures such as the World Wide Web, concepts that are semantically related give rise to substructures that are densely linked. For instance, people interested in databases are going to reference each others' pages. This body of related pages is thus giving rise to a denser body of nodes. Conversely, sets of nodes densely related share common topics and thus correspond to some emergent semantic concept. Thus, in dynamic structures like the World Wide Web, the link topology is fundamentally associated to the semantic lexicon expressed collectively. Most of the techniques developed in Kleinberg and Gibson are concerned with the analysis of static structures. In contrast, the present invention extends techniques to linked data-structures that change with time.

Several scenarios from the intelligence contexts indicate the importance of the detection of concepts from linked data-structure that change with time. In most situations where an international event happens seemingly without warning and surprises the monitoring intelligence agencies, forensic analysis reveals that these agencies had possession of critical information, but that this information never "made it to the top" and was left unutilized. Therefore, a mechanism like the present invention that allows agencies to detect important reports out of the morass of information they routinely process is of prime importance. The present invention includes an intranet-based system of information supporting such automatic detection of concepts.

A further example concerns assisting intelligence agencies in the promotion of the *internal* emergence of critical opinions. The application of the techniques of the present invention to the World Wide Web at large, can help detect and monitor the

emergence of new social movements. The beauty of the approach of the present invention is that it is driven *externally* by the evolution of the World Wide Web, independently of any opinion previously expressed within a monitoring intelligence community.

- Use of the techniques of the present invention is generally justified since
- 5 most “surprising” social events are surprising only because of the inability to read the many dispersed premonitory signals. In actuality, many unrelated individuals notice facts that collectively reinforce each other into a clearer signal. The present invention has a double effect on detection. On the one hand it can “read” the global emergence of signals at levels previously considered to be “below noise level”. On the other hand the present invention
- 10 will help boost the emergence of important detected concepts by publishing such discoveries.

- Existing clustering techniques based on link topology distinguish between *authority nodes* and *hubs*. An authority node is a node that is referred to by many other nodes. For example, the 1905 paper by Einstein is an authority on special relativity. A hub
- 15 node is a node that points to many other nodes. For example, “Yahoo!” is a hub node for the World Wide Web. An authority node is an “important” authority only if it is pointed to by “important” nodes. Conversely, a hub node is an important hub only if points to important authority nodes. This apparently circuitous definition lends itself to a very natural weight diffusion algorithm.

- 20 To illustrate the method, assume that one wants to investigate emergent concepts related to Iraq. One first selects a subpart of the World Wide Web representative of almost all concepts related to Iraq. Specifically, one begins with a seed of (for example) 200 often-referenced sites about Iraq, obtained from a standard search engine like Yahoo! or Alta Vista. Next, the method extends to include all sites that are connected to this initial
- 25 seed. (Actually a bit of pruning is required if too many nodes are connected to that site: think of the site Alta Vista itself!) The graph thus obtained is the graph  $G$  over which the rest of the analysis is conducted.

- Each node  $i$  of  $G$  is allocated two values  $(x_i, y_i)$ :  $x_i$  is its authority value, and  $y_i$  is its hub value. All  $x_i$  and  $y_i$  are initialized to 1. We let  $x^{(0)}$  denote the vector of all initial values  $x_i$  (equal to 1 by definition). Similarly  $y^{(0)}$  is the vector of all initial values  $y_i$ . The
- 30 algorithm proceeds in phases. Each phase  $k$  has two stages. In stage 1, the algorithm updates in parallel all the authority values, transforming the vector  $x^{(k-1)}$  into  $x^{(k)}$ . In stage 2 the algorithm updates in parallel all the hub values, transforming the vector  $y^{(k-1)}$  into  $y^{(k)}$ .
- 35 Specifically, in phase 1, the algorithm updates each  $x_i$  to be the sum of all  $y_j$  for  $j$  pointing to  $i$ . The algorithm normalizes the  $x$ 's so that  $\sum_i x_i^2 = 1$ . Thus, in this update, the authority-

value  $x_i$  of node  $i$  increases if it is referenced by nodes  $j$  with high hub value. In stage 2, the algorithm updates each  $y_j$  to be the sum of all  $x_i$  for  $j$  pointing to  $i$ . The algorithm normalizes the  $y$ 's so that  $\sum_j y_j^2 = 1$ . Thus, in this update, the hub-value  $y_j$  of node  $j$

5 increases if it references nodes  $i$  with high authority value. Thus, hub values and authority values re-enforce each other. One easily establishes that this process converges and that the vectors  $x^{(k)}$  and  $y^{(k)}$  converge to limits  $X$  and  $Y$ . One shows that  $X$  can be directly characterized as being the principal eigenvector of the symmetric matrix  $A^T A$ , where  $A$  is the adjacency matrix of the linked structure; symmetrically,  $Y$  is the principal eigenvector of the matrix  $AA^T$ . Thus, the values  $i$  for which  $X_i$  is "big" are "important" authority nodes. Symmetrically, values  $j$  for which  $Y_j$  is "big" are "important" hub nodes.

A major problem with this technique is the problem of *diffusion*, where, for instance, the original question about Iraq brings sites like Yahoo! or Alta Vista: these sites  
15 are connected to basically everyone and thus appear quite often as important sites in the principal eigenvector. This problem is remedied by considering non-principal eigenvectors: one considers the full spectral decomposition of  $A^T A$  and  $AA^T$ , (not only the principal directions). Each non-principal eigenvector gives rises to a *community* of nodes related by a common concept. The justification is the same as for the principal eigenvector. For every  
20  $n$ , the  $n^{\text{th}}$  eigenvector  $X^{(n)}$  of  $A^T A$  reinforces the  $n^{\text{th}}$  eigenvector of  $Y^{(n)}$  of  $AA^T$ . In more practical terms, the set of sites  $i$  for which  $X_i^{(n)}$  is "high" form a community of authority nodes that reinforce the community of hub nodes  $j$  for which  $Y_j^{(n)}$  is big.

Simulations establish that this technique performs extremely well at  
25 extracting natural concepts from the World Wide Web. It is very robust against variations of the initial seed. The reason is that important hubs and authorities about a subject are by definition reachable from all seed sets of a reasonable size (200 seems to be a reasonable size). In particular, if one considers the World Wide Web to be large in contrast to an intelligence agency's intranet, the technique is very robust against changes of language.  
30 Thus, an initial seed coming from an arabic context will provide very similar results as an initial seed coming from an English context. The reason is that important hubs and authorities are reached from any part of the World Wide Web. That is a big plus for intelligence work! The technique is furthermore computationally quite feasible. The reason  
35 is that the method hinges on the diagonalisation of the matrices  $A^T A$  and  $AA^T$  which are sparse. Accordingly, as is known by those of ordinary skill in the art, there are many efficient iterative methods for performing this task.

The previously described methods apply to the static analysis of a linked topology. The present invention extends these methods to produce a time-varying representation of the concepts of an intelligence intranet or to the World Wide Web. The present invention automatically picks up the emergence of new concepts as they hit a minimal connectivity  
 5 threshold within the intranet or the World Wide Web. It also posts the result of such searches within the intranet of the intelligence agency. Posting the results showing an embryonic emergent new concept will boost its recognition among other participants, if this concept is expressing a genuine social evolution.

The present invention harnesses the diffusion problem. As previously  
 10 explained, the problem is that sharply defined queries will tend to "diffuse" away into more general concepts that have already built a minimal connectivity. To use an image as an example, the diffusion problem is similar to the problem encountered by a distant observer trying to pick at night a neighborhood from among all the lights of a city. This distant  
 15 observer might be able to distinguish a larger neighborhood cluster but would have more difficulty bringing the resolution down to a specific building. The topological approach of the present invention achieves remarkable results by considering large order eigenvectors of the matrices  $A^T A$  and  $A A^T$ . Large order eigenvectors such as the 50<sup>th</sup> non-principal  
 eigenvector do a beautiful job at isolating smaller communities.

FIG. 1 provides a flow diagram of the method 100 for detecting emergent  
 20 concepts from a plurality of sites of the present invention. In step 102, the method 100 for detecting emergent concepts creates an adjacency matrix  $A$ . In step 104, the method 100 computes the transpose  $A^T$  of the adjacency matrix  $A$ . In steps 106 and 108, the method 100 for detecting emergent concepts computes the matrix products  $A^T A$  and  $A A^T$  respectively.

25 Next, in step 110, the method 100 of the present invention selects a value for  $n$ . Using the value for  $n$  selected in step 110, the method 100 for detecting emergent concepts computes the  $n$ th order eigenvector  $X^{(n)}$  of  $A^T A$ . Similarly, using the same value for  $n$  selected in step 110, the method 100 for detecting emergent concepts computes the  $n$ th  
 30 order eigenvector  $Y^{(n)}$  of  $A A^T$ . In step 116, the method 100 determines whether there are any remaining values of  $n$ . If step 116 determines that there are values of  $n$  remaining, then control proceeds to step 110 where the method selects another value for  $n$ . If step 116 determines that there are no values of  $n$  remaining, control proceeds to step 118. In step 118, the method 110 will modify or recreate the adjacency matrix  $A$  to dynamically reflect  
 35 connectivity changes among the sites. Connectivity changes include the addition of connections between sites, the removal of connections between sites and changes in



connection strength. If the adjacency matrix  $A$  is modified control proceeds to step 104 in order to begin computing a new set of eigenvalues.

In an alternate embodiment, the present invention combines the purely topological techniques with a mix of other techniques to control that diffusion. For example, text based techniques allocate a lexical score on communities of nodes containing certain terms. This technique can be used iteratively to refine the graph over which research is performed. Instead of blindly selecting a seed of initial nodes (provided, say, by a standard search engine) and expending it to all the neighboring nodes, this technique selectively constructs the graph by focusing it on the subject at hand. In an alternate embodiment, the present invention also utilizes *latent semantic indexing* as described in Deerwester, Dumais, Landauer, Furnas and Harshman. (1990). *Indexing by latent semantic analysis*. Journal of the American Society for Information Science. 41(1990), 391-407, the contents of which are herein incorporated by reference.

In another alternate embodiment, instead of using a pure adjacency matrix  $A$  whose entries are either 0 or 1 ( $A_{ij} = 1$  if nodes  $i$  and  $j$  are connected), the present invention sets  $A_{ij}$  to different values to account for the strength of the connection. That strength can be evaluated with different filters. For instance,  $A_{ij}$  might be higher if the link was created more recently.

The present invention further includes "time series" analysis tools, where the time series does not track the evolution of scalar values. Instead, the time series tracks the evolution of Web-topological communities. In particular, the growth of new communities can be very instructive and reveal the emergence of new social phenomena.

Further, the detection algorithm of the present invention provides intelligence reports accessible to intelligence participants. The present invention posts these reports on the intranet. The reports themselves become nodes that are linked to the nodes that they have inferred to be linked. Intelligence participants will be able to "answer" these reports by linking to them if they find them worthwhile. Thus, a report would become a catalyst for crystallization of the intelligence, bringing to the fore opinions consensual among a smaller intelligence sub-community.

As mentioned above, the present invention is not restricted to the World Wide Web. Instead, the techniques of the present invention also apply to any linked structure. In particular, one can apply these techniques to monitor the communication patterns of people under surveillance. For instance, one could link two people having communicated within a  $t=24$  hour time window. The spectral techniques described above would allow to pick up communities having tight communication rapport over that period

Standard fraud detection is another application of the link analysis of the present invention. For example, modern computer fraud involves many talented agents, whose individual behavior is apparently normal, but whose collective behavior readily indicates collusion or fraud. Linking these people has been a major intelligence task involving the performance of mostly ad-hoc statistical methods or standard word of mouth. The dynamical linking procedures of the present invention zoom into linking patterns that have been safely ensconced below the detectable detection levels of law-enforcing agencies.

15           The present invention has hardware and software computational requirements because it requires the diagonalization of very large adjacency matrices. On the hardware side, the implementation of such spectral methods require somewhat powerful computing resources. Preferably, the present invention executes on a network of computers as such networks are very powerful and relatively cheap.

On the software side, the present invention requires well-established iterative methods for the singular value decomposition of sparse matrices. As is known by those of ordinary skill in the art, highly optimized code is available to perform this task..

For efficient data processing and archival, it is best to maintain a local copy of the World Wide Web sites over which analysis is to be performed. If not, as mentioned in Kleinberg, the time required to fetch the html-source to construct the base set for the analysis is the greater time bottleneck. Thus, we are thus faced with the standard time/space trade-off.

FIG. 2 discloses a representative computer system 210 in conjunction with which the embodiments of the present invention may be implemented. Computer system 210 may be a personal computer, workstation, or a larger system such as a minicomputer. However, one skilled in the art of computer systems will understand that the present invention is not limited to a particular class or model of computer.

As shown in FIG. 2, representative computer system 210 includes a central 35 processing unit (CPU) 212, a memory unit 214, one or more storage devices 216, an input device 218, an output device 220, and communication interface 222. A system bus 224 is

provided for communications between these elements. Computer system 210 may additionally function through use of an operating system such as Windows, DOS, or UNIX. However, one skilled in the art of computer systems will understand that the present invention is not limited to a particular configuration or operating system.

5           Storage devices 216 may illustratively include one or more floppy or hard disk drives, CD-ROMs, DVDs, or tapes. Input device 218 comprises a keyboard, mouse, microphone, or other similar device. Output device 220 is a computer monitor or any other known computer output device. Communication interface 222 may be a modem, a network interface, or other connection to external electronic devices, such as a serial or parallel port

10           While the above invention has been described with reference to certain preferred embodiments, the scope of the present invention is not limited to these embodiments. One skill in the art may find variations of these preferred embodiments which, nevertheless, fall within the spirit of the present invention, whose scope is defined by the claims set forth below.

15

20

25

30

35

## Claims

1. A method for detecting at least one emergent concept among a plurality of sites comprising the steps of:  
 creating at least one adjacency matrix  $A$ , said adjacency matrix having a plurality of entries,  $A_{ij}$  wherein:  
 $i$  and  $j$  are among said plurality of sites;  
 $A_{ij} = r$  if said sites,  $i, j$  are connected;  
 $A_{ij} = 0$  otherwise; and  
 $r$  is a positive number;  
 computing the transpose matrix  $A^T$  of said adjacency matrix  $A$ ;  
 computing the  $n$ th eigenvector  $X^{(n)}$  of a matrix product of said transpose matrix and said adjacency matrix,  $A^T A$  for determining an authority value of said plurality of sites, wherein  $n$  is a natural number.
2. A method for detecting at least one emergent concept among a plurality of sites as in claim 1 comprising the step of:  
 computing the  $n$ th eigenvector  $Y^{(n)}$  of a matrix product of said adjacency matrix and said transpose matrix,  $A A^T$  for determining a hub value of said plurality of sites.
3. A method for detecting at least one emergent concept among a plurality of sites as in claim 1 wherein said positive number  $r$  represents the strength of said connection between said sites.
4. A method for detecting at least one emergent concept among a plurality of sites as in claim 1 wherein said natural number  $n$  is one.
5. A method for detecting at least one emergent concept among a plurality of sites as in claim 4 wherein said  $n$ th eigenvector  $X^{(n)}$  is a principal eigenvector of said product  $A^T A$ .
6. A method for detecting at least one emergent concept among a plurality of sites as in claim 4 wherein said  $n$ th eigenvector  $Y^{(n)}$  is a principal eigenvector of said product  $A A^T$ .

7. A method for detecting at least one emergent concept among a plurality of sites as in claim 1 wherein said natural number  $n$  is greater than one.
- 5 8. A method for detecting at least one emergent concept among a plurality of sites as in claim 1 wherein said  $n$ th eigenvector  $X^{(n)}$  is a principal eigenvector of said product  $A^T A$ .
9. A method for detecting at least one emergent concept among a plurality of sites as in claim 1 wherein said  $n$ th eigenvector  $X^{(n)}$  is a principal eigenvector of said product  $A^T A$ .
10. Computer executable software code stored on a computer readable medium, the code for detecting at least one emergent concept among a plurality of sites, the code comprising:
- 15 code to create at least one adjacency matrix  $A$ , said adjacency matrix having a plurality of entries,  $A_{ij}$  wherein:
- $i$  and  $j$  are among said plurality of sites;
- $A_{ij} = r$  if said sites,  $i, j$  are connected;
- 20  $A_{ij} = 0$  otherwise; and
- $r$  is a positive number;
- code to compute the transpose matrix  $A^T$  of said adjacency matrix  $A$ ; and
- code to compute the  $n$ th eigenvector  $X^{(n)}$  of a matrix product of said transpose matrix and said adjacency matrix,  $A^T A$  for determining an authority value of said plurality of sites, wherein  $n$  is a natural number.
- 25 11. Computer executable software code stored on a computer readable medium, the code for detecting at least one emergent concept among a plurality of sites as in claim 10, the code further comprising:
- 30 code to compute the  $n$ th eigenvector  $X^{(n)}$  of a matrix product of said adjacency matrix and said transpose matrix,  $A A^T$  for determining a hub value of said plurality of sites.
12. A programmed computer system for detecting at least one emergent concept among a plurality of sites comprising at least one memory having at least one
- 35



## Abstract

5       The present invention relates generally to methods for performing social  
computation. More specifically, the present invention detects emergent concepts from a  
plurality of sites by creating an adjacency matrix representing the connectivity among the  
sites, computing the transpose of the adjacency matrix and computing the nth order  
eigenvalues of the product of the adjacency matrix and the transpose matrix

10

15

20

25

30

35

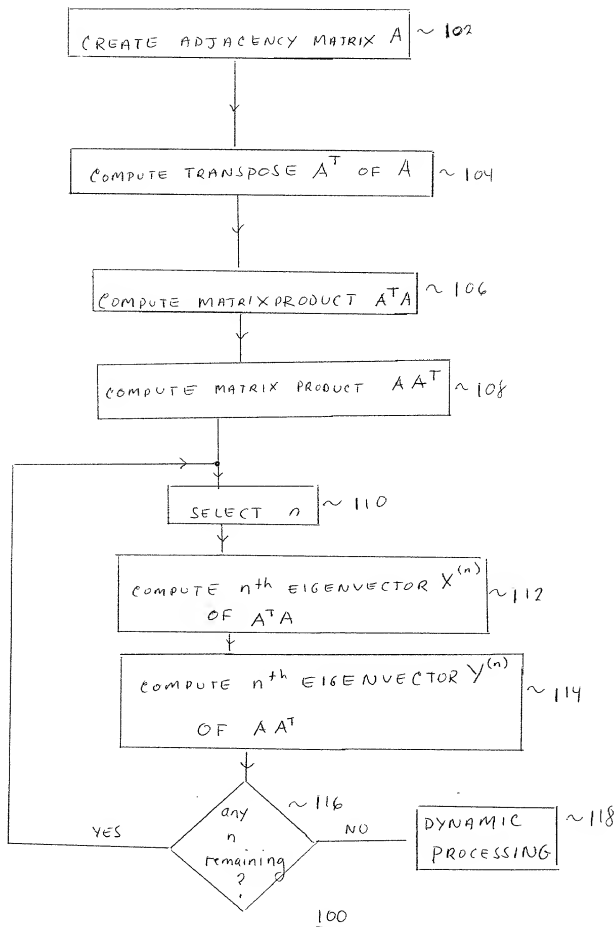
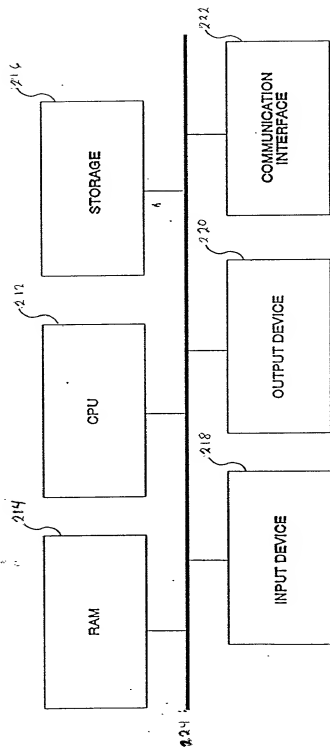


FIG. 1





210

FIG. 2

**DECLARATION  
AND POWER OF ATTORNEY**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below at 201 et seq. underneath my name.

I believe I am the original, first and sole inventor if only one name is listed at 201 below, or an original, first and joint inventor if plural names are listed at 201 et seq. below, of the subject matter which is claimed and for which a patent is sought on the invention entitled

**A METHOD FOR PERFORMING SOCIAL COMPUTATION**

and for which a patent application:

- ☐ is attached hereto and includes amendment(s) filed on 01/11/99 (if applicable)  
☐ was filed in the United States on 01/11/99 as Application No. 09/098,590 (for declaration not accompanying application)  
☐ with amendment(s) filed on 01/11/99 (if applicable)  
☐ was filed as PCT international Application No. 09/098,590 on 01/11/99 and was amended under PCT Article 19 on 01/11/99 (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified application, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119(a)-(d) of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

EARLIEST FOREIGN APPLICATION(S), IF ANY, FILED PRIOR TO THE FILING DATE OF THE APPLICATION			
APPLICATION NUMBER	COUNTRY	DATE OF FILING (day, month, year)	PRIORITY CLAIMED
			YES <input type="checkbox"/> NO <input type="checkbox"/>
			YES <input type="checkbox"/> NO <input type="checkbox"/>

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

APPLICATION NUMBER	FILING DATE
60/098,590	31 August 1998

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code §112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION SERIAL NO.	FILING DATE	STATUS		
		PATENTED	PENDING	ABANDONED

POWER OF ATTORNEY: As a named inventor, I hereby appoint S. Leslie Misrock (Reg. No. 18872), Harry C. Jones, III (Reg. No. 20280), Berj A. Terzian (Reg. No. 20060), Gerald J. Pinfitt (Reg. No. 20823), David Weild, III (Reg. No. 21094), Jonathan A. Marshall (Reg. No. 24614), Barry D. Rein (Reg. No. 22411), Stanton T. Lawrence, III (Reg. No. 25756), Charles E. McKenney (Reg. No. 22795), Philip T. Shannon (Reg. No. 24278), Francis E. Morris (Reg. No. 24615), Charles E. Miller (Reg. No. 24576), Gudon D. Stern (Reg. No. 27469), John J. Laiter, Jr. (Reg. No. 27814), Brian M. Poissant (Reg. No. 28452), Brian D. Coggio (Reg. No. 27624), Rory J. Radding (Reg. No. 28749), Stephen J. Harbulak (Reg. No. 29166), Donald J. Goodell (Reg. No. 19766), James N. Palik (Reg. No. 25510), Thomas E. Friebe (Reg. No. 29258), Laura A. Coruzzi (Reg. No. 30742), Jennifer Gordon (Reg. No. 30753), Jon R. Stark (Reg. No. 30111), Allan A. Fanucci (Reg. No. 30256), Geraldine F. Baldwin (Reg. No. 31232), Victor N. Balancia (Reg. No. 31231), Samuel B. Abrams (Reg. No. 30605), Steven I. Wallach (Reg. No. 35402), Marcia H. Sundten (Reg. No. 30893), Paul J. Zegger (Reg. No. 33821), Edmund R. Barmen (Reg. No. 32110), Bruce J. Barker (Reg. No. 33291), Adriane M. Antler (Reg. No. 32605), Thomas G. Rowan (Reg. No. 34419), James G. Markley (Reg. No. 31636), Thomas D. Kohler (Reg. No. 32797), Scott D. Stimpson (Reg. No. 33607), Gary S. Williams (Reg. No. 31066), William S. Galliani (Reg. No. 33885), Ann L. Gisolfi (Reg. No. 31956), Todd A. Wagner (Reg. No. 35399), Scott B. Familant (Reg. No. 35504), Warren S. Heit (Reg. No. 36828), Kelly D. Talcott (Reg. No. 39582), and Mark A. Farley (Reg. No. 33170) and, all of Pennie & Edmonds LLP, whose addresses are 1155 Avenue of the Americas, New York, New York 10036, 1667 K Street N.W., Washington, DC 20006 and 3300 Hillview Avenue, Palo Alto, CA 94304, and each of them, my attorneys, to prosecute this application, and to transact all business in the Patent and Trademark Office connected therewith.

**SEND CORRESPONDENCE TO:**

PENNIE & EDMONDS LLP  
1667 K Street, N.W.  
Washington, D.C. 20006

**DIRECT TELEPHONE CALLS TO:**  
PENNIE & EDMONDS LLP DOCKETING  
(202) 496-4721

201	FULL NAME OF INVENTOR	LAST NAME SAIAS	FIRST NAME ISSAC	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY Cambridge	STATE OR FOREIGN COUNTRY MASSACHUSETTS		COUNTRY OF CITIZENSHIP USA
	POST OFFICE ADDRESS	STREET 71 Walker Street	CITY Cambridge	STATE OR COUNTRY MA	ZIP CODE 021388
202	FULL NAME OF INVENTOR	LAST NAME	FIRST NAME	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY		COUNTRY OF CITIZENSHIP
	POST OFFICE ADDRESS	STREET	CITY	STATE OR COUNTRY	ZIP CODE
203	FULL NAME OF INVENTOR	LAST NAME	FIRST NAME	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY		COUNTRY OF CITIZENSHIP
	POST OFFICE ADDRESS	STREET	CITY	STATE OR COUNTRY	ZIP CODE
204	FULL NAME OF INVENTOR	LAST NAME	FIRST NAME	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY		COUNTRY OF CITIZENSHIP
	POST OFFICE ADDRESS	STREET	CITY	STATE OR COUNTRY	ZIP CODE
205	FULL NAME OF INVENTOR	LAST NAME	FIRST NAME	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY		COUNTRY OF CITIZENSHIP
	POST OFFICE ADDRESS	STREET	CITY	STATE OR COUNTRY	ZIP CODE
206	FULL NAME OF INVENTOR	LAST NAME	FIRST NAME	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY		COUNTRY OF CITIZENSHIP
	POST OFFICE ADDRESS	STREET	CITY	STATE OR COUNTRY	ZIP CODE

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

SIGNATURE OF INVENTOR 201	SIGNATURE OF INVENTOR 202	SIGNATURE OF INVENTOR 203
DATE	DATE	DATE
SIGNATURE OF INVENTOR 204	SIGNATURE OF INVENTOR 205	SIGNATURE OF INVENTOR 206
DATE	DATE	DATE